FINAL DRAFT

Published as 'Why isn't Stich an ElimiNativist?' in Dominic Murphy & Michael A. Bishop (eds.), *Stich and His Critics*. Wiley-Blackwell, pp. 74-100, 2009.

Why isn't Stich an ElimiNativist about Innate Ideas?*

Fiona Cowie

California Institute of Technology

No, Thursday's out. How about never —
is never good for you?
Robert Mankoff

In 1975, and as usual scratching where it itches, Steve Stich published *Innate Ideas*, a collection of philosophical papers illustrating and analyzing the varied roles of the concept of innateness in philosophical psychology. In the thirty years since then, the idea of the innate has flourished in both popular and academic writings. Not only has it received a fair amount of use (think Chomsky and Pinker; think Fodor; think Herrnstein and Murray; think the *New York Times'* 'Science Times' every other Tuesday), it has also occasioned a good deal of mention, particularly in the last twenty or so years, as philosophers, biologists and psychologists have sought to develop a more precise understanding of what innateness is.

In § 1 of this paper, I briefly survey this large elucidatory literature, starting with Stich's influential account in the 'Introduction' to *Innate Ideas* (Stich 1975). Most of these proposals about innateness are reductive in tenor: they are attempts to specify in scientific terms what it is for some trait – be it a concept, a representation of Universal

Grammar, a heart, or a disposition to fool around – to be innate. As we will see, what emerges from these analytical sorties is not a better grip on innateness, but rather, a

---

* Thanks to Peter Godfrey-Smith, Ben Jeffares and the folks at RSSS, Australian National University, summer 2006, for discussion of this material. And undying gratitude to the editors of this volume, and its victim, for patience well above and indefinitely beyond.

disturbing appreciation of its slipperiness. Innateness has been said to correspond to an array of underlying properties, most of which seem to have little to do with each other. Worse, each proposal is plausible in light of a certain range of cases, but outside that range is beset by counterexamples that cannot be dismissed as deriving merely from peripheral uses of the notion. So while there is a certain 'family resemblance' among the meanings ascribed to 'innate,' there is no analysis of the term that does justice to its multifarious uses. Despite our efforts, innateness insists on remaining a bunch of rather different, if related, things.[1]

Or, perhaps, nothing at all? In §2, I explore the eliminativist option with respect to the innate – what one might call 'elimiNativism.' Very tempting indeed when confronted with the analytical anarchy surrounding 'innateness,' and urged explicitly by behavioral ecologist Patrick Bateson (Bateson 1983; Bateson 1991), neuroscientist Mark Blumberg (Blumberg 2005) and philosopher Paul Griffiths (Griffiths 2002), elimiNativism is the view that no natural kind (or at least, no unified, explanatorily useful property) corresponds to the concept 'innate' and that we should therefore stop talking as if one did. According to the elimiNativist, the concept of innateness does not cut the world along the dotted lines. Instead, it is a pre-scientific or "folk-biological" (Griffiths 2002:76) idea with no role to play in science. Putative explanations in terms of innateness are at best void and at worst misleading, because it encourage fallacies of equivocation (Griffiths 2002:72). Rather than talking of traits that are innate, then, we should explain those traits' development in terms of underlying processes and properties

they

[1] The position I take in this paper is thus a departure from the one I took in Cowie (1999). There, I argued that there are two 'core' senses of 'innate' as used in philosophical psychology. I now doubt that there are just two senses of the term used philosophical psychology, and I am certain that there are many more than that in other contexts.

that are really 'in there' – properties and processes like canalization, generative entrenchment, and a variety of genetic and epigenetic processes and interactions. If we free ourselves of our obsession with innateness, the elimiNativist thinks, we will be in a better position to understand the developmental phenomena that interest us.

I have some sympathy with this line of thought, although as I will argue in §3, there is reason to be cautious in our linguistic and ontological prescriptions. Before I get to this, however, I want to point to a certain oddity in the trajectory of Steve Stich's thoughts about innateness, namely, that *he has never expressed any sympathy with elimiNativism*. Normally not one to pass up the opportunity to hound treasured 'folk' notions (like belief, truth and rationality), Stich has nonetheless remained a champion of the innate – not in the sense that he has explicitly defended any particular nativist thesis, but in the sense that he seems quite committed to the concept's remaining part of the discourse of cognitive science. He has sponsored conferences on innateness, and edited books on innateness (Carruthers, Laurence and Stich, 2005, 2006, 2007), and written influential papers exploring the boundaries of the concept (Stich, 1975, 1978, Ramsey and Stich 1990).

The first section of this paper will motivate the question: Why is this so? Innateness looks ripe for elimination, so why isn't Stich, paragon of conceptual and ontological austerity, an elimiNativist? The rest of the paper will offer an answer to this question on Steve's behalf. I shall argue that Stich is justified in eschewing elimiNativsm for two reasons. First, the arguments for eliminating innateness aren't as strong as they might initially seem (§2). Second, there are powerful pragmatic reasons in favor of keeping the concept current in cognitive science, at least for now (§3).

1.**What is Innateness?**

We begin our survey of the concept of innateness with Stich's 'Introduction' to *Innate Ideas* (Stich 1975). Like Locke, Stich rejects the idea that innateness is a matter of being present at birth: many traits that are thought to be innate (e.g., teeth, perhaps some concepts) emerge much later in development (p.3). He then distinguishes two main senses in which 'innate' is used in ancient and early modern discussions of innate ideas.

The first, deriving from Descartes' analogy between innate ideas and innate diseases, held that a belief (or more generally, a trait) is innate in a person "if, and only if, from the beginning of his life … it has been true of him that if he is or were of the appropriate age (or at the appropriate stage of life) then he has, or in the normal course of events would have, [the trait or] the belief …." (Stich 1975:8) As Stich recognizes, such an account is hostage to one's conception of a 'normal course of events.' Conceive normalcy too broadly, and virtually all traits turn out to be innate; conceive it too narrowly, and none do. While Stich leaves it open that a satisfactory account of 'normal development' could be articulated (i.e., an account such that just the right kinds of things come out innate), others have argued that there is no reason to expect such an account to emerge. (Cf. Cowie 1999; Griffiths 2002; Mameli and Bateson 2005.)

Stich recognizes a second usage of 'innate' in the writings of theorists like Plato and Chomsky. This is the notion that innate ideas (beliefs or concepts) are those that are 'catalyzed' or 'triggered' (pp. 14-15) by experiences that do not seem sufficiently rich or informative to account for the content that the resultant idea ends up having. The same goes for some innate behaviors: the way that a duckling behaves towards its mother seems somehow richer than mere causal commerce with a middle-sized moving body could engender. The idea here is that a trait is innate if its external or experiential causes are in some sense inadequate to explain the trait's existence or properties. The world is merely the 'occasion for' the development of the idea or trait; it doesn't 'construct' the trait.

One difficulty with this account is that it is very restricted in its application: it simply does not apply to many traits standardly held to be innate, like hearts and lungs, learning abilities, or genetic diseases. A second difficulty, and the one Stich emphasizes, is its vagueness: how do we quantify the informational content of beliefs and experiences so as to be able to tell whether the content of a belief outruns that of its trigger? How do we decide whether an external cause (or possibly, a cascade of external causes) is 'constructing' or merely 'occasioning' the development of a behavior? Given that everyone agrees that all traits are partly internally and partly externally caused, how does one assign 'more' of the causative responsibility to one type of factor rather than another? (Cf. Cowie 1999, §1.5.)

Stich leaves his questions about the dispositional and triggering accounts of innateness open; he offers them up, as it were, to the philosophical community as opportunities for further work. However, the philosophical community at the time declined Stich's invitation to elucidate. This is not to say that the issue of innateness went away. On the contrary, Chomsky's provocative claim that linguistic knowledge was innate provoked substantial and spirited debate in the 1970s and 1980s. However, while there was considerable ink expended on the question of whether linguistic (or biological, or mathematical) knowledge is innate, there was very little at the time that explicitly addressed the question of what it is for such knowledge – let alone any other traits – to *be* innate.[2]

From the early 1990's, however, more and more theorists began leaning on the concept of innateness itself. There seemed to be a feeling, among many philosophers interested in the question of how traits develop, that calling some feature or property innate was at best just a gesture towards explaining how it comes into existence. To complete the explanation, one needs to know what innateness is, and how innate traits come to be expressed in the 'finished' organism. Some came at this issue from a more biological point of view; others were more motivated by uses of 'innate' in philosophy, developmental psychology, linguistics, and cognitive science.

1.1 Innateness and cognitive science

Taking the latter group first, I argued (Cowie 1999) that neither of Stich's analyses, even when made more precise, was adequate to account for uses of 'innate' in either the philosophical psychology of the past or in modern cognitive science and linguistics. I argued instead that 'innate' was used in two different ways in these literatures. On the one hand, ideas or bodies of knowledge are called 'innate' when they are acquired as a result of the operation of highly specialized, or domain-specific, mechanisms of learning. What makes an idea or concept innate, on this understanding, is its being got *via* the functioning of a device that is determined or constrained to produce output conforming to that idea. It is in this sense, I argued, that knowledge of Universal Grammar might be innate: we know Universal Grammar to the extent that it is implicit in our knowledge of the grammar of our language, and we know our language in virtue of

[2] A notable exception is Harman (1969a, b).

the operation of a specialized 'language acquisition device' whose job is to produce outputs conforming to UG.

On the other hand, I maintained, 'innate' is sometimes used in a different way in both historical writings and modern cognitive science. At times, the term is used to describe ideas whose acquisition is inexplicable given the explanatory apparatus and concepts of psychology – to express, as I put it, a 'metatheoretic gloom' about our prospects for understanding how the ideas in question are acquired. For instance, when Fodor (e.g., Fodor 1981, 1997, 1998, 2001) says that most concepts are innate, he should be read as meaning that they were acquired by mechanisms the explanation of which is outside the scope of psychology.[3]

A clear difficulty with the first of these accounts of innateness is that because it appeals to mechanisms of learning, it applies only to psychological and behavioral traits. It does not capture what is meant when (say) hearts, or secondary sex characteristics, are said to be innate. A difficulty with the second, metatheoretic (Cowie 1999) or 'primitivist' (Samuels 2002) account is that because it regards a trait as innate if it's inexplicable by psychology, it counts too much as innate: it's not psychology's job to explain why Christopher Reeve was quadriplegic, yet that trait was not innate for him. In addition, primitivism relativizes innateness to our explanatory interests and capacities. This does not seem to be right. The intuition is rather that innateness is a feature that some traits have regardless of our explanatory practices.

One could try to tinker with these accounts in an attempt to fit them to our intuitions. But this just leads to additional problems. For instance, one could count a trait as innate so long as it was acquired by means of *some* specialized developmental

[3] Samuels (2002, 2004) has developed and defended this interpretation, giving it the name 'Primitivism.'

mechanism, that is, a mechanism that had been designed by natural selection to produce that trait reliably at some appropriate point in the course of the organism's development. Ideas got *via* specialized learning mechanisms would count as innate under this construal, as would the appropriate parts and processes of the body and brain. However, other traits, such as Down syndrome or Williams syndrome would not be classified as innate, since the genetic mechanisms giving rise to them have not been selected for that purpose. Also, many traits that are not normally counted as innate would be misclassified. For instance, if mechanisms of psychological and behavioral plasticity have been selected for (cf. Sterelny 2003), then the representations or skills arising out of such learning mechanisms – say, the belief that Dad is dangerous when angry, or the ability to craft a hand axe – would wrongly be classed as innate.

A similar mess results from any attempt to widen the 'gloomy' interpretation of innateness so as to take account of intuitions about innate non-psychological traits. Whatever sciences S one chooses to include in one's expanded definition ("T is innate if it's inexplicable in terms of science S"), one gets some traits misclassified. Worse, the definition fails to deliver univocal answers about the innateness or not of a given trait, since what is taken as primitive for once science is often explicable by another. For instance, neuroscience can explain the development of ocular dominance columns in visual cortex but cognitive psychology takes them to be primitive; psychologists are interested in how we acquire our values but behavioral economists aren't, and so on.

These criticisms do not imply that these are bad analyses of innateness *as used in cognitive science*. What they do show, though, is either that 'innate' as used in cognitive science is a different concept from 'innate' as used in other sciences or by the folk; or else, perhaps, that those whose focus is philosophical psychology are missing a deeper point that talk about innateness is meant to make. After all, intuitions are strong that innateness and the genes have something to do with one another, and yet neither of the 'psychologized' accounts discussed in this section acknowledges this. Perhaps, then, we should look to the term's use in biology, rather than psychology, in out attempt to elucidate its meaning?

1.2 Innateness and biology

Many theorists clearly thought that the answer to this question is 'yes.' For at the same time as cognitive scientists were attempting to understand the meaning of 'innate' in their discipline, a plethora of other accounts, more biological in flavor, began to appear in the literature.

One line of thought, traceable back to Konrad Lorenz, has it that innateness has to do with *genetic determination*. However, as Block (1981), Kitcher (1985), and Mameli and Bateson (2005), among others, have argued, the concept of 'genetic determination' is itself a problematic one, and to explain innateness in terms of it is to trade one analytical pickle for another. For instance, if 'innate' means 'genetically determined' and 'genetically determined' means 'caused by the genes alone' then no trait is innate. But weakening one's notion of 'genetic determination' leads towards errors in the other direction. For instance, if 'genetic determination' means 'genetically influenced,' then everything is innate. Lorenz introduced a notion of 'appropriate causation by the genes' to cash out 'genetic determination.' But this raises the question of what counts as ‘appropriate causation by the genes’ – a question that neither he nor anyone else has ever answered adequately.

A recent paper by Steve Quartz (2003) is illustrative of this difficulty. He articulates five different 'Innateness Proposals,' each specifying a different way to understand the notion of 'appropriate causation by the genes.' A trait is 'strongly innate,' for example, if the neural architecture realizing it is regulated solely by intrinsic processes and mechanisms (genetic and epigenetic); and if T itself develops solely as a result of such intrinsic processes, together with 'merely permissive' environmental interactions[4]; and if T has been 'evaluated' by natural selection. (p. 27). On another hand, a trait T* is innate in the sense of having been 'triggered' if it's just like T, except that instructive as well as permissive environmental interactions are among its causes. (p.28) It's unclear whether any traits are innate by these definitions, and if they are, whether they're the right ones. For it all depends on what's meant by (i) a neural structure's being regulated solely by intrinsic factors and (ii) the notion of a permissive (vs. instructive) environmental interaction. But we will not pursue these issues here. For suppose that Quartz is right, and there are five kinds of genetic causation, each covering some but not all prototypically innate traits. That in itself is almost sufficient to make the point I'm aiming at in this section, namely, that 'innate' means so much to so many that it's in danger of meaning nothing at all.

A second major biological approach to innateness cashes out that concept in terms of *genetic coding* or *genetic information*: a trait is innate if it is coded in the genes, or if

[4] A ‘permissive’ environmental interaction is one that affects development by virtue of its mere presence or absence. This is to be contrasted with an ‘instructive’ interaction, where features of the environment affect development by virtue of their statistical structure or informational content. (Quartz 2003:30) Thus, the ingestion of food is a merely permissive environmental input, while being told that Elizabeth I bathed once a month whether she needed it or not is instructive.

all the information required for the trait's development is encoded in the genes, the environment playing merely a supportive, non-informational role (cf. Maynard Smith, 1993; Fodor, 2001, Bates et al., 1998; Fodor 1998, 2001; Quartz 2004). Such accounts must explain the notions of genetic encoding or information they invoke. Predictably, given that these are semantic notions, problems arise (Godfrey-Smith 1999, 2000a, 2000b; Griffiths and Gray 1994, 1997, 2001; Griffiths 2001).

There are two main approaches to understanding genetic coding or information, and hence innateness, on the present interpretation. First, correlational accounts say that T is genetically encoded (innate) when certain genetic properties – say, a deletion at a certain locus – correlate with T. (E.g., Sterelny and Kitcher, 1988.) But if correlation between As and Bs sufficed for A's encoding B, then environments would code for traits, too. However, they don't: being raised in an English-speaking environment correlates pretty well with – but doesn't encode -- being an English speaker.[5] Second, selectional accounts of genetic coding or information say that a gene, G, carries information about T if G has been selected for its role in producing T. (E.g., Maynard Smith, 2000, Sterelny and Griffiths, 1999, Sterelny 2000). But diseases like Huntington's chorea are innate, and are said by experts to be 'genetic' or 'genetically encoded,' yet they have not been subject to any selection at all, so far as we know, for they present only after reproduction has occurred.[6] In addition, Godfrey-Smith has argued that if genes code for or carry information about any traits at all, it's most plausible to think that the relevant properties

[5] Such objections motivate some developmental systems theorists' skepticism about the theoretical usefulness of the gene-environment distinction, e.g., Griffiths and Gray, 1997.

[6] Now that genetic testing can reveal the presence of some genetic diseases before symptoms appear, selection against them may become quite strong, as individual reproductive choices are influenced by knowledge hitherto unavailable.

are properties of proteins or other factors, not the high-level phenotypical properties that are the concern of nativists (Godfrey-Smith 2004, see also Samuels 2004).[7]

A third biological approach to innateness is to focus on the link between innateness and *developmental invariance*. Both Bill Wimsatt and Andre Ariew, for instance, explore the idea that innate traits are *developmentally invariant:* they appear very reliably, and are very hard – and often dangerous – to interfere with. Wimsatt argued that innateness should be understood in terms of *generative entrenchment*. (E.g., (Wimsatt 1986; Wimsatt 1999).) On his view, a trait is innate if it has come under strong stabilizing selection due to its playing a role in the development of many other adaptive features, and if this stabilizing selection has resulted in the processes responsible for the trait's development having been made more robust (i.e., harder to interfere with). Ariew's proposal defined innateness in terms of the related but distinct notion of *canalization*. (Ariew 1996, 1999.) On this view, a trait is innate if its development is highly canalized, that is, buffered against a wide range of initial conditions and environmental variables – if it develops pretty much 'come what may.'

These accounts seem to capture, albeit in different ways, the innateness of (i) bodily parts and processes; (ii) basic neural and psychological machinery; (iii) various behaviors and behavioral dispositions; and (iv) species-typical traits (like menstruating or growing body hair) which reliably occur later in development. However, they suffer from two serious flaws. One derives from the fact that learned beliefs and behaviors can be entrenched and/or canalized in exactly Wimsatt and Ariew's senses. Yet it's often said to be a *sine qua non* innateness that nothing learned is innate. Our acquisition of the

[7] Indeed, the occurrence of 'alternative splicing,' whereby genes may conform themselves in different ways to give rise to different protein products in different circumstances, may problematize even the idea that genes code for proteins, if coding is understood as a correlational notion.

vocabulary of a natural language is a case in point. It is buffered against what certainly seems to be a wide range of environmental variations: except for cases where the linguistic environment is radically impoverished, *all* children, *whatever* their circumstances, learn the vocabulary of their language. Further, it's plausible that our ability to learn the phonological forms and meanings of words has been selected for and made more robust because of its vital role in the acquisition of linguistic competence, itself vital in our acquisition of numerous other adaptive features and behaviors. If so, then the mechanisms that allow us to acquire a vocabulary, whatever they are, would have become both canalized in Ariew's sense, and entrenched in Wimsatt's sense. Acquiring a vocabulary is buffered, and it is entrenched; yet vocabularies are learned, hence paradigmatically not innate.[8]

Another potential problem with both Ariew and Wimsatt's accounts is that they make innateness a matter of degree. For Ariew, traits seem to be more or less innate depending on the scope of the environmental buffering they are blessed with. For Wimsatt, traits would seem to be more or less innate depending on (i) the number of 'downstream' developmental processes that depend on them and/or (ii) their degree of entrenchment or robustness. This by itself is not necessarily a problem for these analyses: surely the idea that some traits are more innate than others can be accommodated by the folk. What is a problem, though, is the fact that the deliverances of these analyses are highly sensitive to one's choice of a certain *range* of environments (in Ariew's case) or particular *notion* of robustness (in Wimsatt's case). Consider an

[8] It's consistent with this claim that other aspects of our knowledge of language are innate. Indeed, the argument can be made more generally: bracket off whatever parts of language you take to be inborn; the rest still has to be learned, and because of the value of having a language, it's plausible that that learning process will be both buffered and entrenched in our species.

example due to Mameli and Bateson (2005: 20): the development of normal hands and feet. Hands and feet are buffered against interference in a wide range of environments *so long as* the relevant range does not include environments in which thalidomide circulates in the placenta. If, however, the range does include such environments, then hands and feet are not buffered in that range. Likewise, one might ask, in what sense is getting hands and feet a robust developmental process if it can be interfered with so easily and drastically *via* the ingestion of thalidomide during pregnancy? Why choose one range of environments rather than another? Why choose a particular notion of robustness? It won't do simply to pick the ranges and ideas of robustness that 'give the right answer,' for these will differ from case to case. We need a principled answer to these questions.

A fourth biological notion that has been appealed to in attempting to clarify innateness is *heritability*: a trait is innate if it is highly heritable. (Cf., e.g., Herrnstein and Murray 1994). However, this approach is bankrupt. First, traits that are universal in the population – like having a head – have a heritability of 0, so would not be deemed innate. Secondly, learned traits – like being able to read – have high heritability in many populations. Which brings us to a final problem for heritability-based accounts of innateness: since heritability is a population measure, heritability estimates vary with one's choice of population. Thus, among children in California, having freckles on one's nose is highly heritable: basically, you get them if you're Anglo, but otherwise, not. Among children in Australia, by contrast, freckles are much less heritable, even among Anglos. For having freckles depends on exposure to the sun, which in turn is affected by whether your school has (as many do) a policy that sunhats must be worn when playing outside. To define innateness in terms of heritability is to inherit this counterintuitive population-relativity.

Finally, in this survey of attempts to 'biologize' innateness, we come to two related ideas: innateness as *species typicality* and innateness as *adaptatedness*. Again, neither is satisfactory, either alone or in conjunction. Some species-typical adaptations are learned or otherwise culturally transmitted: tool-making in *homo erectus;* some birds' characteristic songs; wearing clothes in modern *homo sapiens*. And some innate traits (being a haemophiliac, being homozygous for the sickle cell gene) are neither species-typical nor adaptive. Being an adaptation is neither necessary nor sufficient for being innate, and the same goes for being a species-typical trait.

1.3 The Inner and the Outer

A third major line of thinking about innateness, and the last we will be covering here, is less overtly biological in tone. It focuses on the intuition that innate traits in some sense 'come from inside,' as opposed to 'coming from the environment.' However, this intuition proves difficult to nail down. Since all traits are produced *via* interactions between inner and outer, one cannot simply say, for instance, that innate traits are ones that are not caused by the environment, or that innate traits are the ones that are produced by "interactions internal to the organism" (Elman et al. 1996:23).

But as we have seen already, in discussing Quartz's (2003) attempts to come to grips with innateness, there are many ways that genes and environments collaborate to produce phenotypes and it's not obvious which of these ways should be held to produce the innate ones. Moreover, as Mameli and Bateson argue in some detail, analyses in the 'inner vs. outer' vein founder when they confront particular cases. (Mameli and Bateson 2005:8-9) For instance, consider the notion that a trait is innate if it is not alterable by environmental factors. This gets being a queen bee right: being a queen is not innate because whether a female becomes a queen or not depends on the amount of royal jelly she ingests. However, it gets having arms wrong: that trait is innate, despite its being environmentally alterable (say, by thalidomide, or by a Congolese machete). Or consider the idea that a trait is innate if it's unalterable in *normal* environments: that would cope with the case, just discussed, of having arms. However, as Mameli and Bateson rightly emphasize and as we have seen above (§1.1), appeals to normalcy carry their own problems. If the relevant notion is statistical normality, then what is innate becomes unacceptably hostage to changing fortune and fashion. But if it's evolutionary normalcy that is at issue, then we must (i) nail down the period of interest for the evolution of the trait in question, and (ii) determine what normal conditions at that time were, and (iii) decide whether the trait was alterable in those conditions or not. None of these is an easy task, of course, but even given an account of normal conditions in the Pleistocene, say, it is plausible that we would find many behaviors and properties that we would not want to count as innate despite their being unalterable in those conditions: eating only minimally processed foods, or having a short life expectancy are examples.[9]

A related approach resting on the inner vs. outer distinction contrasts the innate with the learned. Thus, Fodor (1975, 1981) equates innate concepts with those that are unlearned. But, first, this approach makes innateness hostage to our conception of what is to count as learning – a notoriously thorny issue in itself. Moreover, it does not

[9] Of course, this argument depends on adopting a particular reading of 'alterable.' I won't explore this issue here, except to note that the fact that alterability is a modal notion reveals another dimension of vagueness that dogs this analysis (as well as providing a potentially fertile source of counterexamples to it).

generalize beyond the mental: there are many unlearned traits – having a tattoo, for example – that are not innate. Mameli and Bateson (p. 15) explore the possibility of invoking the broader notion of *phenotypical plasticity* (rather than the narrowly cognitive notion of learning) in this context. For instance, they consider the idea that a trait might be innate iff it does not result from mechanisms designed to produce plasticity, i.e., "mechanisms adapted to produce different traits in response to different environmental conditions." (Mameli and Bateson 2005:15). Again, though, this analysis encounters counterexamples and problems. For instance, there is a sense in which the genetic mechanisms responsible for sex determination in humans are adapted for plasticity: add androgens to the placental mix and you get a male rather than a female. Yet surely one's sex, of all things, is innate? Moreover, a *failure* to be produced by such mechanisms is not sufficient for innateness. While the analysis gets such things as menstruation and body hair right (they are innate because the mechanisms producing them are not adapted for plasticity), it gets loads of other traits wrong: thalidomide-affected limbs, acquired sociopathy, Christopher Reeves' quadriplegia: not a result of mechanisms adapted for plasticity, but not innate either.

1.4 Innateness: a too many splendour'd thing.

In § 1 so far, we have examined the following ways of understanding what it means to say that a trait, T, is innate:

1. T is innate if T is present at birth
2. T is innate if T emerges in the normal course of development
3. T is innate if T's external or experiential causes are inadequate to explain its existence or properties.

4. T is innate if T is acquired as a result of the operation of a highly specialized, or domain-specific, mechanism of learning.
   4a. T is innate if T was acquired by means of a developmental mechanism designed by natural selection reliably to produce T at the appropriate point in the organism's development.
5. T is innate if the processes responsible for T's development are inexplicable given the explanatory apparatus and concepts of psychology. ('Primitivism')
   5a. T is innate (by the lights of some science, S) if the processes responsible for T's development are inexplicable given the explanatory apparatus and concepts of S.
6. T is innate if T is genetically determined.
   6a. T is innate if T is caused by the genes alone.
   6b. T is innate if T is genetically influenced.
   6c. T is innate if T is appropriately caused by the genes.
   6d. T is innate if (i) T has been selected for and (ii) the architecture of the part(s) of the brain responsible for the acquisition of T developed under the control of genetic factors together with merely 'permissive' environmental factors.[10]
7. T is innate if T is genetically encoded.
   7a. T is innate if all the information required for T's development is encoded in the genes.
8. T is innate if T is generatively entrenched.
9. T is innate if T is highly canalized.
10. T is innate if T highly heritable.
11. T is innate if T is species typical.
12. T is innate if T is an adaptation.
13. T is innate if T is "produced by internal causes."
14. T is innate if T is not caused by the environment.
15. T is innate if T is not environmentally alterable (i.e., if changes in the environment cannot produce alternative phenotypes T*).

[10] A 'permissive' environmental interaction is one that affects development by virtue of its mere presence. This is to be contrasted with an 'instructive' interaction, which affects development by virtue of its statistical structure or informational content. ((Quartz 2003:30)

15a. T is innate if T is not alterable in normal environments (i.e., if alternative phenotypes T* cannot occur in normal environments).

16. T is innate if T is unlearned.

16a. T is innate if does not result from mechanisms designed to produce plasticity.

Yet not one of these sixteen (or twenty-five, depending on how you count) accounts appears adequate to capture everything that we mean when we claim that a trait is innate. My conclusion echoes that of (Mameli and Bateson 2005): the concept of innateness is fragmented and confused, and every analysis ever proposed misclassifies central exemplars of the concept it purports to elucidate.

2. The Case for ElimiNativism

Given the mess that the folk concept of innateness is patently in, it's tempting to ask whether we oughtn't to try to do without it. In this section, we will look at whether there is reason to pursue this course.

After distinguishing two different kinds of eliminativist project (ontological and linguistic) and explaining how they are related, I turn to the arguments for elimiNativism. There are three kinds of argument used to support an eliminativist position – what I call the **'Aren't Any**,' **'Doesn't Work**,' and **'It's a Mess**' arguments. In §§2.2-2.4, I illustrate these arguments with examples taken from other areas of philosophy, and look at how they might apply to the issue of innateness. In §2.5, I suggest that taken together, these arguments raise a conundrum for theorists interested in development, which I then explore in §3.

2.1 Two Eliminativist Projects and Three Kinds of Eliminativism

I distinguish two different kinds of eliminativist project: ontological and linguistic (or conceptual). The ontological project asks whether a given type of object exists or whether a certain property is instantiated. As it pertains to innateness, the ontological question is whether any trait out there in the world is in fact innate. The linguistic (or conceptual) project asks whether a given theoretical term (or a given concept) should be used or eschewed: does the term 'innate' have any substantive role to play in philosophical or cognitive scientific theorizing? (Henceforth, for ease of exposition, I will talk of eliminating terms and eliminating concepts interchangeably. The eliminativist arguments I will consider apply to both, *mutatis mutandis.*)

These projects are related, but distinct. If a certain object or property does not exist, then there may be reason to stop using terms that purport to refer to it (in certain contexts, at least). But equally, there may be reasons to continue with that usage: centers of gravity do not exist, but there are many contexts in which it's not just OK but arguably necessary to think and talk – and design bridges, or figure skate, or play pool – as if they did. Electrons are not just particles, and atoms are not like tiny solar systems, but when doing chemistry, one often talks as if they were. Conversely, the fact that a certain term is unfitted to play an explanatory role in any scientific or philosophical theory may be reason to think that the class or property the term purports to pick out does not exist; but equally it may not. 'Socks', 'Hummers,' 'dirty jokes' and 'herbs' are terms that no reputable philosophical or scientific theory needs or wants; yet all those things are really out there, nonetheless.

In this paper, I will talk of eliminating 'innateness' when it's the linguistic question at issue; when it's the ontological question we are concerned with, I will talk of eliminating innateness (no quotes).

Orthogonal to the distinction just made is one between the three kinds of eliminativist argument to be found in the philosophical literature. The first is what I call the 'there aren't any' argument for eliminativism, or **'Aren't Any Eliminativism'** for short. Aimed squarely at the ontological issue, it takes an analysis of a certain concept and argues that nothing in the world does (or, stronger, could) satisfy that analysis. This is the eliminativism pursued, for instance, by Hume with respect to causation, by Mackie (1981) with respect to moral properties, and more recently by Stich with respect to the propositional attitudes. Aren't Any eliminativists may or may not advocate linguistic eliminativism as well: Stich (1983) thinks we can do without 'belief' and the like, but compare Hume's (sometime) views about 'causation' or Mackie's attitude towards moral talk.

A second eliminativist strategy is **'Doesn't Work Eliminativism'**. Aiming primarily at the linguistic question, the Doesn't Work eliminativist contends that the concept of interest is enmeshed in a bankrupt theory: one that is false, explanatorily inadequate and part of a thoroughly 'degenerate' research program. On this view, the offending theory and its component concepts should be expunged from our thought, and its linguistic correlates should be replaced in our discourse by other, better, theories and terms. The ontological question in this case remains open. Because false theories can succeed in referring to real entities and properties (and here arise issues in the philosophy of language that I will not explore), the Doesn't Work eliminativist argument, even if successful, does not entail any position on the ontological question.

A final kind of eliminativism is **'It's a Mess Eliminativism,'** again aimed in the first instance at the linguistic or conceptual question. But rather than taking an analysis of a concept and arguing that nothing satisfies it, as does the Aren't Any eliminativist, and rather than arguing that the term or concept is embedded in a useless and dispensable theory (like the Doesn't Work eliminativist), the It's a Mess eliminativist argues that the concept in question *has* no determinate analysis, and that it is too vague or confused to be used profitably in a certain context. We see this form of argument in, for instance, Woodward (2003) with respect to the concept of a law of nature, and in Godfrey-Smith (1999, 2000a) with regard to 'genetic information.'

Frequently, It's a Mess eliminativists take their arguments for linguistic elimination to support ontological elimination too: given that it's so unclear what the concept is a concept *of* the existence of its object(s) must be dubious as well. However, the mere fact that a concept is unanalyzable can't by itself entail any sort of eliminativism: *most* concepts are unanalyzable! So the It's a Mess eliminativist's argument must include a claim to the effect that there is something about the uses to which the concept is put, or the practical consequences of its unanalyzability, that cause unusual difficulties in this case. If, for instance, the eliminativist claim were just that concept C is unsuitable for use in some science, or in metaphysics or philosophy of science, it may not follow that there are no Cs. We've seen this with socks and herbs. 'Memory' is another case in point, as David Chalmers (pers. comm.) has pointed out to me. 'Memory' (unqualified) refers to a plethora of mental states and processes (long term memory, short term memory, procedural memory, somatic memory, etc.). As such, it lacks the precision that neuroscientists and philosophers dealing with these phenomena require, and for the purposes of theorizing, most experts on memory are eliminativists about 'memory.' None, however, are eliminativists with respect to memory!

With these brief sketches in hand, let us see in more detail how the various strategies work, and how they might be wielded against innateness.

2.2 Aren't Any Eliminativism and its Bearing on ElimiNativism

The Aren't Any eliminativist takes an analysis of some concept or term, like 'good' or 'bad' or 'belief' or 'true' or 'rational', and argues that there is nothing in the world that corresponds to that analysis, that is, that there is no such thing as goodness, or badness, or beliefs, or truth. Confronted with this situation, one should probably replace talk of the good or beliefs or truth with other, more accurate ways of talking. But if we continue to use these dubious terms and concepts, we must be aware that when we think or talk in those ways, we often speak and think falsely.[11]

A good example of Aren't Any Eliminativism is Steve Stich's (1983) argument for the view that folk-psychological concepts like belief "*ought not* to play any significant role in a science aimed at explaining human cognition and behavior" (1983:5, emphasis Stich's). Stich's argument begins with an analysis of the notion of belief. When we ascribe the belief *that p* to someone, he maintains, we attribute to her a certain relation towards a token sentence, stored somehow in her brain, which means *that p.*

[11] Or maybe even meaninglessly, depending on your semantic theory. Note that whatever his semantic theory, the Aren't Any eliminativist can't hold that *all* talk involving such terms is false or meaningless: 'There are no beliefs' is a statement that Stich and Churchland both embrace! Thanks to Frank Jackson for this point.

Two mental sentence tokens (hence two belief tokens) are of the same type in virtue of their contents. Thus, these mental sentences, Stich argues, are individuated ‘widely’ (or ‘semantically’), not ‘narrowly’ (‘syntactically’ or ‘functionally’).

However, Stich contends, being of the same belief-type is not a matter of having *the same* content. Instead, the folk individuate beliefs in terms of *similarity of content.* This raises a problem. Since similarity judgments are notoriously context-sensitive and vague, a taxonomy of mental states based on them will be vague and context-sensitive too. Worse, since the folk judge similarity of content relative to their own contents, our taxonomy will be observer-relative and limited in application to thinking beings very like the folk.

But the taxonomy of mental states used in cognitive science won’t be vague, context-sensitive, observer-relative, or limited in application! Its taxonomy will be precise, objective, and applicable to any thinking thing. For the purposes of science, Stich argues, the folk’s talk of beliefs individuated in terms of content will be replaced with talk of different kinds of entities – ‘beliefs*’ – which are individuated in terms of their syntactic (narrow, functional, non-semantic) properties. Propositional attitudes* – not the propositional attitudes of the folk – will be what are quantified over in the sciences of the mind. So according to the scientists, there’ll be nothing in our heads that has the properties (i.e., the semantic properties) that we ascribe to beliefs, etc. In which case, Stich intimates, the beliefs, desires, hopes and fears of the folk must be eliminated: “[i]f our science is inconsistent with the folk precepts that define who and what we are, then we are in for rough times. *One or the other will have to go.”* (Stich 1983:10, emphasis added.)

Returning to innateness, it is immediately clear that the Aren't Any strategy is unable to support ElimiNativism – and I suspect that this is one reason why Stich is not an elimiNativist. The Aren't Any argument against propositional attitudes worked because Stich was able to establish that propositional attitudes are *essentially* contentful: if anything is to count as a belief, it *has* to be the kind of thing that is individuated in terms of its content. Given the plausibility of this essentialist claim about beliefs, Stich is able to argue that since a scientific taxonomy of the mental is very unlikely to end up individuating mental states in terms of content, the ontology of folk psychology will not be empirically vindicated.

Of course, there are many things to be said about all this [REF to someone else in the volume?] Our concern, however, is with the applicability of this form of eliminativist argument to innateness, and it is very clear that *it does not apply at all*. For what is striking about the plethora of analyses of innateness surveyed in the previous section is that there seems to be *nothing* that is essential to the folk concept 'innate'!

But if there's nothing that's essential to innateness, then the Aren't Any eliminativist's strategy cannot even begin to get off the ground. And thus, returning to the question posed in §1, Stich is quite right not to have applied the Aren't Any strategy that worked so well in the case of propositional attitudes to the question of innateness.

But there are other eliminativist arguments to consider: should Stich have pursued one of them?

2.3 Doesn't Work Eliminativism and the Innate.

The Doesn't Work eliminativist argues that the concept or term of interest is embedded in a false, unproductive, or otherwise scientifically useless theory. For that reason, he maintains, we should cease using that concept. A classic example of this eliminativist strategy is Paul Churchland's assault on the propositional attitudes of folk psychology (Churchland 1981). He argues, first, that folk psychology is an empirical theory, on a par (in terms of its epistemic status) with other empirical theories. For instance, it makes generalizations that quantify over a variety of theoretical entities, such as beliefs and desires, and it is sensitive in a variety of ways to evidence. Since folk psychology is an empirical theory, he points out, it "might (really) be false" (72). Which, he thinks, it is. First, Churchland argues that although it does enjoy some explanatory successes (you and I succeed in meeting up at the Impeachment Rally in West LA because each believes the other will be there, each wants to see the other, and each takes appropriate steps, based on her further beliefs and desires, to arrive on the right day at the right time at the Federal Building in Westwood), folk psychology nonetheless is weighed down with numerous serious explanatory failures. It has nothing to say about mental illness, creativity, the origins of intelligence differences, the nature and functions of sleep, visual perception and illusions, memory, or learning. Because folk psychology offers so little insight about so many matters ostensibly within its domain, Churchland argues, it is "*at best* a highly superficial theory, a partial and unpenetrating gloss on a deeper and more complex reality." (74) As to its history, "the story is one of retreat, infertility and decadence." (74). Its domain has contracted; primitive people used folk psychology to explain the workings of the world quite generally. Its understanding of those things left in its domain (viz., our minds) "[has] not advanced sensibly in two or three thousand years," (74) this despite the "backlog" of anomalies and explanatory failures that should have spurred its refinement. In addition, folk psychology appears not to cohere with other human and biological sciences: "its intentional categories stand magnificently alone, without visible prospect of reduction [to other sciences]." (75)

In sum, folk psychology, together with its classification of mental states as beliefs, desires, sensations etc., is a highly limited theory, rife with explanatory failures and anomalies, that has been stagnant and in retreat for millennia. Just as folk psychology turned out to be false of the weather (clouds do not move because they want to), it will likely turn out to be false of us and our behavior. This idea is supported by the fact that despite years of effort, there is no prospect of reducing its categories to those of 'respectable' sciences: the mind-body problem resists solution. Folk psychology is ripe for elimination – and the entities, properties and processes it postulates are likewise ready to disappear.

Can this type of argument be run in the case of innateness? It is true that the innate has resisted reduction to the categories of some or other science – the motley collection of analyses in §1 is testament enough to that. But is it true that the theory has stagnated, or that its domain is in retreat? I don't think so. To be sure, there are things that nativists cannot explain, and there are limits to the explanations they offer, but the charge of degeneracy can hardly be leveled at the developmental sciences that are currently making use of the idea of innateness. On the contrary, they have been hugely productive, especially over the last few decades. The search for the innate has inspired philosophers, psychologists, biologists and other scientists not only to talk about what 'innate' means, but to investigate the development of so-called innate traits in extraordinary depth, leading to a vast explosion of knowledge about how the processes of development unfold in us and other species.

The case of language acquisition is very much to the point. Chomsky's claim that universal grammar is innate has inspired several generations of smart young people to devote their lives to explaining how language is acquired. This was not even considered a problem in the 18th, 19th or early 20th centuries, and although the behaviorists in psychology had begun to sniff around the issue by the late 1950s, Chomsky's claim in the early '60s that language is innate focused people's attention on this phenomenon in an unprecedented way, and led to the development of a huge body of new knowledge about how languages are learned, stored, used and lost. In light of developments like these – and they are paralleled in all sorts of other areas, such as the study of concept acquisition in psychology and philosophy, the study of birdsongs and various 'instinctual' behaviors in ethology, and the study of visual perception, face recognition, causal reasoning and mathematical abilities in psychology and neuroscience. There simply is nothing degenerate about the nativist's research program: the Doesn't Work eliminativist's argument has no purchase here.

2.3 **It's a Mess Eliminativism and its Implications for ElimiNativism**

However, the It's a Mess eliminativist's argument remains to be examined. §1 constituted its first premiss: 'innate' is a term that stubbornly resists analysis. The second premiss asserts that the concept is *so* hopelessly muddled that it warrants elimination, *regardless* of any fruits borne by its deployment. Griffiths (1997, 2002) and Bateson (1983, 1991) advance arguments along these lines.

Griffiths (2002, 71) identifies three 'core features' as characteristic of the innate:

(i) **Developmental fixity**: innate traits are hard to change;

(ii) **Species nature**: what is innate is somehow related to what it is to be a member of a certain species;

(iii) **Intended outcome**: innate traits are non-accidental; they are somehow *meant* to be there.

These three properties, he argues, pretty much capture what we pretheoretically take innateness to be. Moreover, he argued, each of them has a 'respectable' and related successor concept in modern biological science: the successor to developmental fixity is canalization; the descendant of species nature is species typicality; and the notion of an adaptation provides a scientifically decent gloss on what it might be for some outcome to be intended.

However, he continues, if that is what innateness is, then that notion is in deep trouble. For these properties frequently come apart in actual cases. For instance, bilateral symmetry in vertebrates is highly canalized and typical of the phylum, but is not an adaptation, being rather the result of developmental constraints. A Live Oak's generously spreading crown is species-typical and adaptive, but that growth pattern is not canalized: in high altitudes, the Live Oak's profile is sparse and scrubby. An Australian Aborigine's dark skin is adaptive and canalized, but not species-typical. Griffiths argues that because the properties clustering together under 'innateness' need not co-occur, continuing use of the term encourages fallacies of equivocation, such that a trait's evolutionary origin (say) is inferred from its species typicality, or its immutability is inferred from its being an adaptation. If we mean developmental fixity, species typicality or whatever, we should just say so.

From their consideration of the various proposed analyses of innateness, Mameli and Bateson (Mameli and Bateson 2005) identify a larger cluster of 'core' features of innateness, which they call 'innateness properties' or 'i-properties.' These are (p. 26):

(1) reliably appearing in a particular stage in the life cycle
(2) being such that environmental manipulations capable of producing an alternative trait are evolutionarily abnormal
(3) not being produced by a mechanism adapted to map different environmental conditions onto different phenotypes and, at the same time, not produce by the impact of evolutionarily abnormal environmental factors
(4) being generatively entrenched
(5) being developmentally environmentally canalized (i.e., development is buffered against a wide range of environmental changes)
(6) being post-developmentally environmentally canalized (i.e., once the trait has developed, it is stable under a wide range of environmental variation)
(7) being species-typical
(8) being a standard Darwinian adaptation

They point out that whether or not these 'i-properties' are highly correlated or not "is an empirical question, and an important one." (p. 27). Answers to it allow us to answer a range of other, important empirical questions (e.g. "How often and in what circumstances does natural selection result in developmental or post-developmental canalization?" (p.27)). In addition, they would also have implications for elimiNativism: "If the i-properties cluster, then it is appropriate to say that the folk concept [of innateness] does track a real biological phenomenon: … it tracks what we can call the 'innateness syndrome.'" (30) In that case, they argue, the concept of innateness would be "scientifically useful in that it could play a positive role in theory development and in the discovery of generalizations." (30) However, if the i-properties are not highly correlated, "we shouldn't try to use the notion of innateness while doing science…[in that case] the various debates that have been framed in terms of innateness are better dealt with by referring to each of the i-properties individually." (30-31)

2.4 **Something of a Paradox**.

The arguments discussed in §2.3, as well as the rather motley character of Bateson and Mameli's list of 'i-properties' itself, however, suggest that the i-properties are not highly correlated. By Bateson and Mameli's lights, as by Griffiths', then, the concept 'innate' should be eliminated.

But to accept this conclusion is to confront a somewhat paradoxical situation: 'innate' is a concept that is inadequate in numerous ways, failing to carve nature at its joints, meaning different things to different people, and thereby generating fallacious reasoning, confusion and an inordinate amount of analytical despair. Yet at the same time, it seems to be a linchpin of a vital and exciting set of explanatory research programs and projects in the sciences of the mind (§2.3). As analytical philosophers, it appears, we can't live with the concept; yet as cognitive scientists, it appears just as plainly, we can't shoot it.

What is to be done?

I suggest: nothing.

In §3, I argue, first, that it is quite common in science for investigators to continue using concepts that everyone knows to be inadequate – ambiguous, vague, incomplete, perhaps even incoherent. I maintain, secondly, that this practice seems de facto to be part and parcel of the development of new, better ways of conceiving the world. I urge, third, that there are philosophical (not just historical) reasons to expect this to be the case. And thus I conclude that it's a bad idea on both pragmatic and philosophical grounds to eliminate bad concepts -- like 'innateness' -- until better ones are found.

3. Good Uses for Bad Concepts

3.1 'Gene'

In two recent papers (2006, forthcoming), Karola Stotz and Paul Griffiths describe the development of the concept 'gene.' The term 'gene' was introduced by Johannsen (1911), but the origins of the corresponding concept are much more obscure. In large part, this is because there wasn't and still isn't a single gene concept in use in biology, or even in genetics itself. In the early days, most geneticists were instrumentalists about genes or 'factors': they were defined functionally, as the elements (possibly fictional) that were transmitted from parent to offspring and explained observed patterns of inheritance. Each Mendelian trait, or 'character' was associated with a single ‘factor,’ and traits that did not follow a Mendelian inheritance pattern were held to be produced by two or more genes.

But little attention was paid to the question of what genes were. Even Thomas Hunt Morgan, whose group established that genes are arranged linearly on chromosomes, was content to regard them as purely hypothetical entities, mere fictions used in calculating the observable results of hybridization experiments, “because at the level at which genetic experiments lie, it does not make the slightest difference” (quoted G&S, ‘instrumental gene’ section, para 1). However, Herman J. Muller, a student of Morgan, was dissatisfied with the instrumental notion of ‘gene’ and set out to discover genes’ physical basis. According to him, genes were physical entities that could replicate and mutate, and that played a role in the construction of phenotypes. Whether they corresponded to the classical geneticist’s factors was for Muller unimportant. Thus gene-properties that for the classical concept had been essential, because definitional, were now treated as accidental and open to question.

The advent of the field of biochemistry, between the two world wars, led to further refinements of the gene concept. For instance, now genes had to be the kinds of molecules that could create other, very specific molecules, with highly constrained conformations: genes were individuated one-to-one with their specific enzymatic products. It was thought that genes were proteins, because the structures of proteins, like those of enzymes, were tightly constrained. This ‘one gene-one enzyme’ view led early evidence that genes were composed of DNA to be ignored, as DNA did not seem to have the complex structure the hypothesis required. It was not until the 1940s, when new experimental techniques and assumptions were introduced into biochemistry by physicists such as Max Delbruck, that it was accepted that genes were made of DNA.

In the early 1950s, Richard Goldschmidt had pointed out that the geneticist's instrumentalist conception of a gene (and related concepts like 'mutation') was incompatible with the material concept emerging in biochemistry. He urged that the biochemist's conception was the correct one, famously opining that there were no genes, at least in the classical sense. Genetic eliminativism!

After the structure of DNA was elucidated by Watson and Crick, further changes in the gene concept occurred. For instance, Muller had held that genes were by definition the unit of recombination, which was held to occur only when alleles from the two different arms of the chromosome come together during meiosis. Seymour Benzer, however, established that recombination could occur within a gene, as parts of the gene changed position. This led to the dropping of the requirement that genes be what are recombined in (sexual) reproduction, and the foregrounding of genes' functional role: genes became the segments of DNA that produce the molecules of RNA that produce a specific protein (or, later, other product).

At the same time as the molecular gene concept was evolving, so too was the geneticist's more instrumentalist conception. By the 1970s, geneticists had abandoned their forebears' insistence that genes were what mutated and recombined; all that remained of the classical conception was genes as contributors to phenotypes. Griffiths and Stotz argue that at this point, the two conceptions were compatible and that both sides (so to speak) were more or less in agreement about what genes are. And although they go on to show that this halcyon period was brief, with ‘gene’ again coming unstuck (and remaining so today), Griffith and Stotz conclude that this story represents

> …a highly successful example of the research strategy of identifying a functional role, searching for the mechanism that fulfills that role at a lower level of analysis, and using knowledge of that mechanism to refine understanding of function at the original (in this case phenotypic) level of analysis. (p. xxx)

And indeed it does. However I wish to draw a slightly different moral from their tale. The philosopher-historian telling a hundred years of history in thirty pages can only gesture at the turmoil, breast-beating (and, occasionally, despair) that accompanied this conceptual evolution as it was actually going on. At the time when Goldschmidt declared that there are no genes, it appeared to those in the field not just that the geneticists’ concept was never going to square with that of the biochemists’, but – and importantly – that the biochemical concept itself was hopelessly muddled, empirically unjustified, perhaps at times even empirically invalidated. Yet it appears to us *just obvious* that the field would have made a gross mistake had it followed Goldschmidt’s eliminativist advice and given up on genes – certainly the things, but even the word or concept. Many new discoveries and techniques, not to mention better concepts, were born of the jostling between the apparently incompatible notions of gene-as-replicator, gene-as-protein, gene-as-locus-of-mutation, gene-as-producer-of-phenotypical-characters, etc. And while the counterfactuals are, of course, speculative, it is hard not to believe that had any of these been jettisoned prematurely, the nature of the relations between them would not have been clarified, but rather obscured further, to the likely detriment of both biological research and our understanding of the world.

Good things can come from ‘bad’ concepts.

3.2 'Principle' or ‘Element’

In this section, we look at another case illustrating how science makes eminently productive use of concepts that everyone knows to be inadequate. The historian Ursula Klein (2005) traces the development of organic chemistry from pre-1750 botanical practices up until the late 19th century, when modern classifications and nomenclature emerged. I argue that this account demonstrates that particularly when the structure of some natural phenomenon or domain is genuinely up for grabs, it can be necessary – in the strongest, albeit pragmatic sense – to continue using concepts that one knows are ripe for elimination in order to design and carry out the very experiments that will reveal the true nature of that domain.

Prior to 1750, European chemists accepted the ancient distinction between the three natural ‘kingdoms’ of plants (‘vegetables’), animals, and minerals. (Klein, 2005: 268-9 In their attempts to classify vegetable substances further, they drew on the practices of those involved in the non-scientific manipulation of plants and their derivatives, for instance, apothecaries, dyers, brewers, and traders. Numerous lists classifying plant materials were published around this time, each differing significantly from the others as to the placement of particular substances. Nonetheless, according to Klein, classification was organized around three (maybe four) main categories: (i) *natural origin* (species, country or city, type (e.g., Balsam of Peru)); (ii) *mode of extraction* (substances made by nature (‘simples’) vs. substances made by human art -- ‘composita’ (mixtures) and ‘preaparata’ (simple chemical extractions)); (iii) *Observable or sensible* *properties* (waxes, gums, leaves, etc. and their properties – e.g., balsams have a strong aromatic smell, syrupy consistency, yellow-brown color, and minimal or no solubility in water); and, less commonly, (iv) *Practical uses* (e.g., substances that were used in brewing and distilling were grouped together, as were those used in the production of wines and spirits, as were those used as dyes.

Notably *not* among these classificatory principles is *what the substance is made of*. This is not because no-one was interested in the composition of plant and other natural substances. On the contrary, the search for the alchemist's 'ultimate principles,' the underlying natures that caused the perceived properties of natural substances, was most definitely afoot at this time. Nor is the chemists' lack of taxonomic interest in constituent structure due to no-one's having thought to base classification on constituency. Georg Ernst Stahl, for instance, had proposed a taxonomy that classed substances based on their composition and modes of chemical combination in 1730. (Klein, p.269, n.12.)

Rather, the search for constituents had little or no impact on the classificatory project because "chemists did not consider knowledge of composition a reliable precondition to order all of the kinds of materials they were dealing with." (269) And as we now know, they were right! For the constituents that were being sought (and 'found') at this period were the 'ultimate principles' postulated in the alchemical philosophy, such as phlegma, spirit, oil, salt, and earth. Since they don't exist, and since substances have always refused to behave as if they did, taxonomists were right to ignore them.

In their taxonomic endeavors, in other words, chemists were behaving as if they were eliminativists about principles or constituents – and about 'principles' and ‘constituents.’ Those terms and properties were not relevant to the scientific task at hand, and were not used or adverted to. But it would have been a mistake for these theorists to have elevated their decision to pursue one taxonomic course rather than another into a matter of eliminativist principle. Any of the arguments we examined in § 2 – Aren’t Any, Doesn’t Work, It’s a Mess – could have been run on the notion of a principle or constituent to justify eliminativism, ontological and linguistic. But the abandonment of all talk and thought of principles in chemistry from 1750 onwards would have been a disaster, as the rest of Klein’s story indicates.

For after 1750, constituency was gradually foregrounded as the foundation for vegetable taxonomy. Spurred by the discovery of substances’ ‘proximate principles’ or ‘compound components,’ and spurring in turn numerous further developments in chemical practice and theory, the search for the elemental structure of organic substances culminated 150 years later in Dumas’ 8-volume *Traité de chimie appliqué aux arts* (1835), widely regarded as the first work of organic chemistry as we know it today. The interplay between changing conceptions of substances’ principles or elements, new experimental techniques and instruments, new representational techniques, and developing theories about the organization of nature that made this magnificent achievement possible simply could not have happened had some hypothetical eliminativist about principles or elements ‘won’ in the 1750s.

The first development relevant to the emergence of a constituent taxonomy was a proliferation, around mid-century, of a variety of new chemical techniques. In addition to distillation, decoction and combustion, chemists began to stress the use of mechanical means (such as grinding), low-temperature (‘wet’) distillation, and various solvents as being important means for the chemical analysis of plant substances. (283) These new analytical methods served in the first instance to reveal that the old taxonomy (outlined above) was inadequate. For instance, the old distinction between simples and *praeparata* was undermined when analysis showed that some simple substances (like saps and resins) were more similar to various *praeparata* (like essential oils) than to other simples (say, leaves). Thus, the French chemist Venel (along with many others) classified saps and essential oils together as being among the proximate principles or compound constituents of plants. (287)

A second vital development was conceptual. Between the 1750s and 1790s, chemists recognized plants as *organized* living beings, and began replacing the three 'kingdoms' of the previous centuries with a bipartite division of nature into organized or 'organic' beings (plants and animals), on the one hand, and those that are not so organized (minerals), on the other. This reclassification was no simple merging of the animal and vegetable kingdoms. For instance, Rouelle placed some plant-derived substances, such as balsams, resins, and gums on the mineral side of the divide, and in 1766, Macquer attributed a vegetable origin to the former 'mineral,' bitumen. Macquer also proposed a criterion of the organic in terms of constituency, rather than origin: both animal and plant bodies contained "an inflammable, fatty or oily substance" (291) whereas minerals did not, and by 1790, it was recognized that certain mineral salts were not among the principles of plants, even though they could be extracted from plants. The old taxonomy was breaking down, but no new orthodoxy had yet emerged.

A third precondition for the rise of a new classificatory schema was the importation into organic chemistry of discoveries about inorganic substances. In 1789, Lavoisier, together with a number of other chemists, had published the *Méthode de nomenclature chimique*. It codified the trend that was emerging more slowly in plant chemistry, explicitly excluding from taxonomic consideration such properties as place in the three kingdoms, sensible properties, manner of production or extraction, and practical application. Instead, it advocated the classification metal oxides, acids and salts based on their composition. The new, 'elemental' chemistry was vastly more explanatory than the old, for Lavoisier had coupled a 'theory of affinity' to his theory of composition. According to the affinity theory, the various elemental constituents of inorganic substances bore mutual relations of affinity with one another, and were in equilibrium only at ordinary temperatures. This explained why upon heating, the various component parts were released from the bonds of affinity, and made manifest. The notion of affinity also explained why different 'proximate principles' were observed under different conditions.

By the early 1800s, the community of chemists, impressed with the new chemistry's successes in the inorganic realm, recognized that this kind of elemental approach was, in principle, the way to go in classifying organic substances as well. Lavoisier himself had suggested applying the elemental approach to organic substances, proposing in the *Méthode* that "the true constituent elements of vegetables are hydrogen, oxygen and charcoal" (L. quoted K, p.300.) However, no serious attempt to analyze plant substances in terms of these basic elements was made until well after the turn of the century. (Lavoisier had abandoned the search for the principles of plants soon after the *Méthode* was published.) Instead, chemists continued to try to develop taxonomies that were recognizable descendants of the older, more practical and phenomenological approach. Some substances got put in different classes, to be sure, and the classes themselves were played with and renamed. But further developments were necessary before the 'elemental' approach to taxonomizing organic substances could finally be taken.

A major impediment to such a step, according to Klein, was a lack of conceptual tools for understanding how the various elemental substances could combine to produce the variety of observed organic substances. While chemists around 1800 knew that organic substances were made of hydrogen, oxygen and carbon, and while they wanted to taxonomize them in these terms, they could not "because of the absence of a theory of definite quantitative composition that would have allowed the clear identification and demarcation of the different kinds of plant substances." (314) J.J. Berzelius recognized the problem acutely. His experiments had shown that wildly different substances – for instance, sugar and gum – contained the same proportions by weight of the elemental substances:

Sugar:	6.802% hydrogen
	44.115% carbon
	49.083% oxygen
Gum:	6.792% hydrogen
	41.752% carbon
	51.456% oxygen (Klein 2005:314)

How, then, were the two to be distinguished? Berzelius recognized that by assuming that different elemental substances had 'atoms' of different characteristic weights, he could get clearly different accounts of their composition:

Gum: 12 O + 13 C + 24 H

Sugar: 10 O + 12 C + 21 H

This – also the origin of modern chemical nomenclature – was a major breakthrough. However, a further theoretical problem, noted but not addressed by Berzelius, remained. For it seemed clear that the mere numbers of the different 'atoms' could not account for the observed differences among organic substances: some substances that differed only a little in the number of atoms they contained differed a lot in their observed properties, and conversely.

This problem was solved by Dumas and Boullay. In 1828, they argued that not only was the *number* of atoms of each element important to a substance's identity, but so too was *how the atoms were put together*. They proposed that the substances that they had grouped together – namely, two kinds of sugar, ethal (a substance isolated from fat), alcohol, and ethers – were not directly composed of the elements, but nonetheless shared the same 'binary constitution' (K:316): they were all "compounds of bicarbonated hydrogen" and the differences between them were explained in terms of differences in the other, non-bicarbonated component. (316) Dumas and Boullay's insight is still celebrated as the beginnings of organic chemistry as we know it today. (318)

By 1835, substances were no longer being classified according to superficial similarities (sensible or chemical) or in terms of their uses or origins. Instead, as Dumas urged in his 8-volume *Traité de chimie appliqué aux arts* they were classified "from a purely chemical standpoint" (Dumas, quoted in Klein, 2005:318) – that is, in terms of "their atomic weights and…the rational formulae which represent their nature and their properties" (p. 319). Even substances made in the laboratory were to be treated in this way, so long as their constituency – being compounds of carbon -- marked them as organic. And substances formerly thought to be emblematic of plant chemistry – such as wood fibers, fibers, corn starches – were no longer of interest to plant chemists. Carbon chemistry had begun, and with it came a whole new ontology for organic chemistry.

More good things from (really) terrible concepts.

4.**Against Premature Elimination**

In §3, with the examples of 'gene' and 'principle' (what we now call 'element'), I have tried to show that conceptual confusion is not unheard of in science; nativists are not alone in having only a dim and confused idea of what it is they are talking about. Further, as I tried to show in § 3.1, the clashing of incompatible concepts can be highly fruitful of research: the battle of the instrumental vs. the material gene led workers both in genetics and in biochemistry to examine their own concepts more closely than they would perhaps otherwise have done, to the benefit of both disciplines. And finally, as I've tried to bring out in § 3.2, scientists often have no choice but to keep working with concepts that they know to be inadequate. It's simply not possible to design, conduct and interpret the very experiments one needs in order to show the way forward without using some concepts – and if the bad old concepts are the only ones available, well: it's put up or shut up. Organic chemists could not have simply tossed out the ultimate principles of the alchemists, or the proximate principles of the pre-1750s taxonomies, *until they had something else to put in their place.* And even when they *knew* what the correct principles were – oxygen, hydrogen, and carbon – they *still* were unable to theorize in terms of them until a wealth of other knowledge and techniques had been developed *using precisely the incorrect principles they were aiming to replace*.

Modern discussions of innate traits look a lot like 18th century discussions of vegetable materials. Different taxonomies of developmental processes abound (canalized vs. not; entrenched vs. not; learned vs. not; genetically determined vs. not). It seems that our talk of innateness is functioning to delineate in a general way what the topic of inquiry is, just as the early 'ultimate' and 'proximate' principles, and the taxonomies that went along with them pointed in the direction of the truly organic.

And just as late 18th century chemists knew that their taxonomic principles were the wrong ones, yet couldn't burn the boat they were fishing from, we can know that the innate/not innate distinction is not quite the right one to make, yet keep right on making it nonetheless. These are concepts that 'glom onto' important classes of closely related phenomena, even though they don't get the boundaries of those classes exactly right, and even though they don't properly taxonomize the kinds of phenomena that are in the family they refer to. Socrates said, more or less, that you can't learn something if you don't already know it. I say: you can't investigate something if you don't have a way of thinking about it. The concept of innateness, I contend, enables us to think about developmental phenomena that we don't yet fully comprehend. Throw that concept away, and we lose that ability; lose the ability to think about those phenomena, and you lose the ability to investigate them.

With these points in mind, we can return to the question that heads this paper: why isn't Stich an elimiNativist? We saw in §2 that the 'Aren't Any' argument doesn't run, because there's nothing that's essential to innateness; that the 'Doesn't Work' argument doesn't work, because the nativist research program is flourishing; and that the 'It's a Mess' argument poses a conundrum. On the one hand, 'innate' really *is* a mess, and as philosophers, appalled at the sight of a concept so déshabillé, we long to draw a veil. On the other hand, though, there is that siren research program to contend with. What to do?

On the basis of the historical examples discussed in §3, I have suggested that in this kind of case, philosophical prudery should yield to a more permissive pragmatism. Good things can and sometimes do come from thoroughly misbehaved concepts. Premature elimination can be disastrous, obstructive to progress and destructive of thought itself. As long as 'innate' appears to be swinging along alright, it's wiser not to interfere; when it's time for the concept to retire, it'll be clear – or at any rate, clearer than it is right now.

My brief in this paper was to be critical. And indeed, I began with the idea of showing how Steve's Quinean, scorched-earth nerve had failed him in respect of the innate. As it turned out, however, I discovered that there were good reasons not to be elimiNativist. At least, they seem good to me, and I expect that they probably seem good enough to Steve as well, himself a vigorous shoot from the Peircean root. If I'm right about this, though, it strikes me as an interesting further question why pragmatic considerations should incline him to hang on to the innate, at the same time as he is quite happy to jettison the intentional. I suspect that another of his Peircean biases – the pro-science, anti-philosophy one this time – is at the bottom of that.

But that's a topic for another time. Maybe Thursday.

References

Bateson, P. (1983). Genes, environment and the development of behavior. Genes, Development and Learning. P. S. a. T. Halliday. Oxford, Blackwells**:** 52-81.

Bateson, P. (1991). Are there principals of behavioral development? The development and integration of behaviour. P. Bateson. Cambridge, Cambridge University Press**:** 19-39.

Block, N. (1981). Introduction: what is innateness? Readings in the Philosophy of Psychology. N. Block. New York, Methuen. **2:** 279-281.

Blumberg, M. S. (2005). Basic Instinct: the genesis of behavior. New York, Thunder's Mouth Press.

Carruthers, P., S. Laurence, et al. (2005). The innate mind: structure and contents. New York, Oxford University Press.

Carruthers, P., S. Laurence, et al., Eds. (forthcoming). The Innate Mind: Culture and Cognition. New York, Oxford University Press.

Churchland, P. M. (1981). "Eliminative Materialism and the Propositional Attitudes." Journal of Philosophy **78**(2): 67-90.

Cowie, F. (1999). What's within? nativism reconsidered. New York, Oxford University Press.

Elman, J. L. (1996). Rethinking innateness: a connectionist perspective on development. Cambridge, Mass., MIT Press.

Fodor, J. A. (2001). "Doing without *What's Within*." 2001, from http://host.uniroma3.it/progetti/kant/field/cowiesymp.htm.

Fodor, J. A., T. G. Bever, et al. (1974). The Psychology of Language. New York, McGraw Hill.

Griffiths, P. E. (2002). "What is innateness?" The Monist **85**(1): 70-85.

Griffiths, Paul E. and Karola Stotz (2006). "Genes in the Postgenomic Era?". Theoretical Medicine and Bioethics, 27(6): 499-521.

Griffiths, Paul E. and Karola Stotz (In Press). "Gene", in Michael Ruse and David Hull (eds.), Cambridge Companion to Philosophy of Biology, Cambridge: Cambridge University Press, xxx-xxx.

Harman, G. H. (1969a). Linguistic Competence and Empiricism. Language and Philosophy: A Symposium. S. Hook. New York and London, NYU Press and University of London Press**:** 143-151.

Harman, G. H. (1969b). "Psychological Aspects of the Theory of Syntax." Journal of Philosophy **64**: 75-87.

Kitcher, P. (1985). Vaulting ambition: sociobiology and the quest for human nature. Cambridge, Mass., MIT Press.

Klein, U. (2005). 'Shifting ontologies, changing classifications: plant materials from 1700 to 1830.' Stud. Hist. Phil. Sci. 36 (2005) 261–329

Mameli, M. and P. Bateson (2005). "Innateness and the Sciences." Biology and Philosophy**?**

Quartz, S. R. (2003). "Innateness and the brain." Biology & Philosophy **18**(1): 13-40.

Ramsey, W. and S. Stich (1990). "Connectionism and 3 Levels of Nativism." Synthese **82**(2): 177-205.

Samuels, R. (2002). "Nativism in cognitive science." Mind & Language **17**(3): 233-265.

Samuels, R. (2004). "Innateness in cognitive science." Trends in Cognitive Sciences **8**(3): 136-141.

Sterelny, K. (2003). Thought in a hostile world: the evolution of human cognition. Malden, MA, Blackwell.

Stich, S. P. (1975). Introduction: the idea of innateness. Innate Ideas. S. P. Stich. Berkeley, University Of California Press**:** bibl.

Stich, S. P. (1983). From folk psychology to cognitive science: the case against belief. Cambridge, Mass., MIT Press.

Stich, S. P. (1990). The fragmentation of reason: preface to a pragmatic theory of cognitive evaluation. Cambridge, Mass., MIT Press.

Wimsatt, W. C. (1986). Developmental constraints, generative entrenchment, and the innate-acquired distinction. Integrating Scientific Disciplines. W. Bechtel, Martinus-Nijhoff.

Wimsatt, W. C. (1999). Generativity, entrenchment, evolution adn innateness: philosophy, evolutionary biology, and conceptual foundations of science. Where biology meets psychology: philosophical essays. V. G. Hardcastle. Cambridge, Mass., MIT Press.